Supporting Information

Constantin de Magny *et al.* 10.1073/pnas.0809654105

SI Text

Seasonal Patterns. During the 9-year time period considered in this study, standardized monthly mean time-series of cholera for Kolkata showed a unimodal distribution, with an annual progressive increase in the number of cases beginning in March, a major peak in the fall, and highest variability in September (supporting information (SI) Fig. S1). However, ongoing studies indicate that the annual distribution of cholera cases in Kolkata may have changed over time, with the modern period exhibiting bimodality (K. Rajendran, personal communication). However, our complete time series analysis did not show this bimodality. The distribution of cholera incidence annually in Matlab is clearly bimodal, with a minor peak in the spring and a major peak in the winter (Fig. S1). Much more variability in CHL was observed for Kolkata between May and November, with a peak in August. However, the distribution was bimodal in Matlab, with less variability in the spring and greater variability between July and December (Fig. S1). The distribution of standardized SST data for the two areas also revealed a bimodal seasonal pattern. For Kolkata, SST was higher in June and during September–October (Fig. S1), while for Matlab the SST was higher in May and October (Fig. S1). The standardized rainfall distribution was similar at both locations, with the distribution approximately unimodal, peaking in July, but with a detectable bimodality in Kolkata.

Contribution of Individual Factors. Quantitative relationships between observed cholera cases and significant factors in the models, i.e., $CHL_{ano(t)}$ and $PRECIP_{ano(t)}$ for Kolkata and $CHL_{ano(t-1)}$ for Matlab are illustrated on Fig. S2.

Incidence. In Matlab, the total population in the surveillance area in 1990 was 203,490 and was 225,750 in 2006. In 17 years, the population increased by about 22,260 (+11%), *i.e.*, a 0.65% increase per year. To evaluate any potential effect of using incidence instead of confirmed cholera cases in our analysis, we converted the confirmed cholera cases time series between 1997 to 2006 to incidence per 100,000. We linearly interpolated the population size from the two censuses. We obtained the same model structure for the two locations, as described in the manuscript. The coefficients changed slightly due to the different range in response variable.

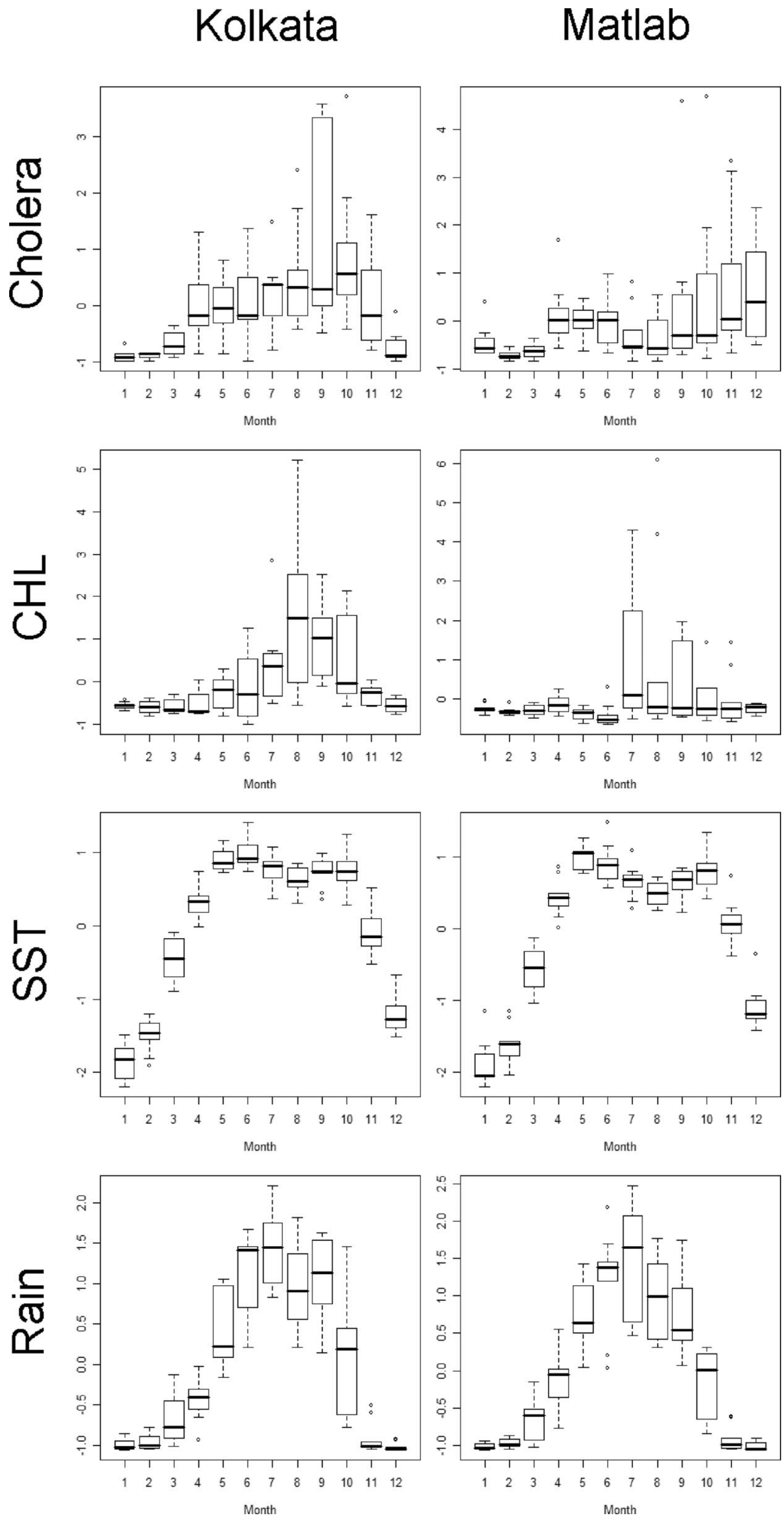

Fig. S1. Mean seasonal distribution (box-plots) of number of cholera cases and environmental and climate data (Sep-1997 to Dec-2006). From left to right, the panels correspond to locations in Kolkata (India) and Matlab (Bangladesh), respectively. From top to bottom, the panels correspond to number of cholera cases, chlorophyll concentration (CHL), sea surface temperature (SST), and rainfall, respectively. *x*-axis represents month and *y*-axis the standardized value.

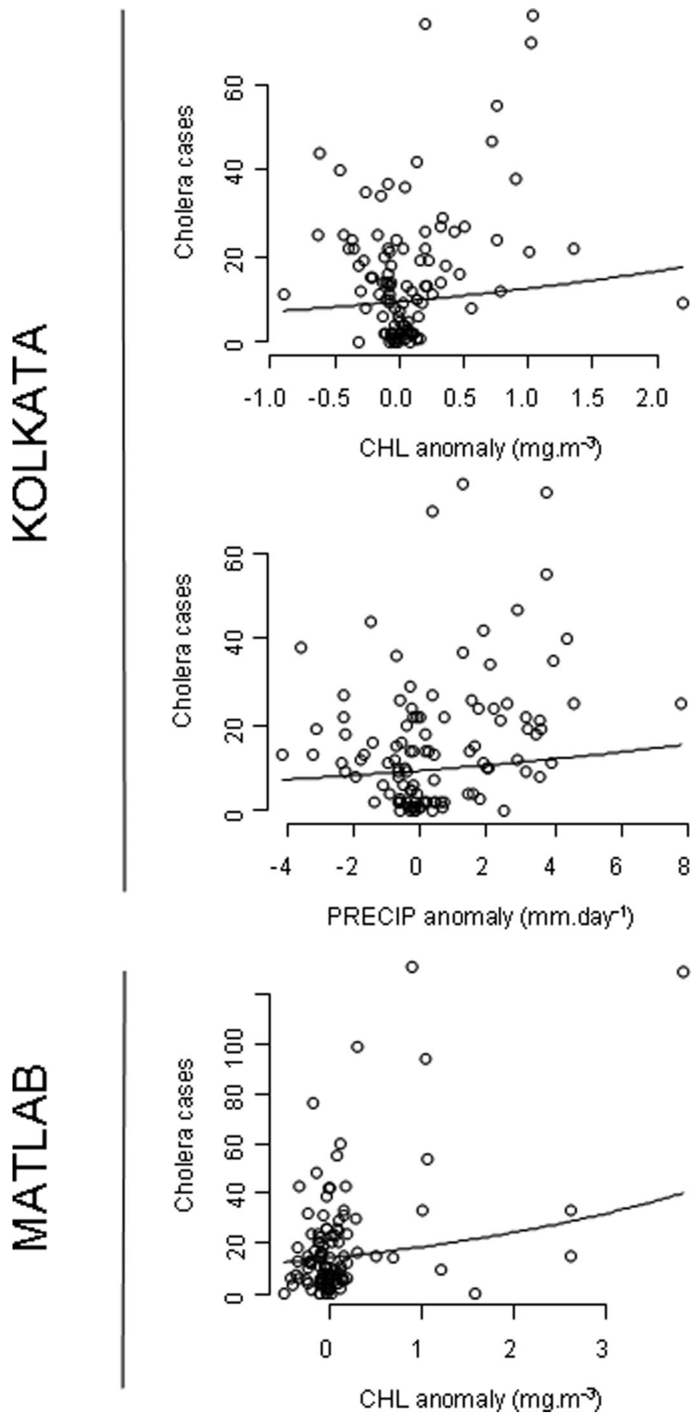

Fig. S2. Scatter plots of the observed cholera cases and significant factors retained in the models. (*Top*) Kolkata observed cholera cases vs. $CHL_{ano(t)}$. The black line corresponds to the regression curve of the contribution of $CHL_{ano(t)}$ with all other variables set to their mean. (*Middle*) Kolkata observed cholera cases vs. $PRECIP_{ano(t)}$. The black line corresponds to the regression curve of the contribution of $PRECIP_{ano(t)}$ with all other variables set to their mean. (*Bottom*) Matlab observed cholera cases vs. $CHL_{ano(t-1)}$. The black line corresponds to the regression curve of the contribution of $CHL_{ano(t-1)}$ with all other variables set to their mean.